

The Second Workshop on SciSoc Agents & LLMs: Agentic AI for Scientific and Societal Advances

Jiayuan Ding
Hippocratic AI
Palo Alto, CA, USA
jiayuan@hippocraticai.com

Wei Jin
Emory University
Atlanta, GA, USA
wei.jin@emory.edu

Carl Yang
Emory University
Atlanta, GA, USA
j.carlyang@emory.edu

Yixin Wang
Stanford University
Palo Alto, CA, USA
yixinwang@stanford.edu

Chanyoung Park
Korea Advanced Institute of Science
and Technology
Daejeon, Republic of Korea
cy.park@kaist.ac.kr

Xiangliang Zhang
University of Notre Dame
Notre Dame, IN, USA
xzhang33@nd.edu

Jimeng Sun
University of Illinois
Urbana-Champaign
Urbana-Champaign, IL, USA
jimeng@illinois.edu

Danai Koutra
University of Michigan
Ann Arbor, MI, USA
dkoutra@umich.edu

B. Aditya Prakash
Georgia Institute of Technology
Atlanta, GA, USA
badityap@cc.gatech.edu

Yan Liu
University of Southern California
Los Angeles, CA, USA
yanliu.cs@usc.edu

Abstract

The Second Workshop on SciSoc Agents & LLMs: Agentic AI for Scientific and Societal Advances aims to explore the emerging paradigm of agentic artificial intelligence, where Large Language Models (LLMs) are combined with autonomous agents, tools, memory, and multi-step reasoning to drive scientific discovery and address critical societal challenges. As LLMs continue to advance in capability, their integration into agent-based systems enables more interactive, goal-driven, and adaptive workflows that go beyond passive text generation. In particular, Agentic AI systems have demonstrated strong potential in orchestrating complex scientific pipelines, reasoning over large and heterogeneous datasets, interacting with external environments, and supporting decision-making in domains such as healthcare, environmental science, education, public policy, and social sciences. By bringing together researchers and practitioners from AI, data mining, science, and society, this workshop aims to foster a comprehensive understanding of how LLMs and agent-based architectures jointly redefine traditional research methodologies. Participants will explore novel agentic frameworks, multi-agent collaboration, and real-world deployments that leverage LLM-powered agents to accelerate scientific progress and generate meaningful societal impact.

Keywords

Agentic AI, AI Agents, Large Language Models, Multi-Agent Systems, AI for Science, AI for Social Good

ACM Reference Format:

Jiayuan Ding, Wei Jin, Carl Yang, Yixin Wang, Chanyoung Park, Xiangliang Zhang, Jimeng Sun, Danai Koutra, B. Aditya Prakash, and Yan Liu. 2026. The Second Workshop on SciSoc Agents & LLMs: Agentic AI for Scientific and Societal Advances. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26)*, August 09–13, 2026, Jeju Island, Republic of Korea. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3770855.3818253>

1 Workshop Overview

Topics. We invite submissions from diverse fields at the nexus of AI, science, and society, including but not limited to healthcare, environmental science, education, public policy, social science, chemistry, and biology. We particularly encourage contributions that demonstrate the practical applications of Large Language Models (LLMs) [1, 10–12] and autonomous agents [4, 7, 8] in addressing real-world challenges. Relevant topics, **focused on scientific and societal applications**, include but are not limited to the following list:

- Pre-training and fine-tuning of foundation models for agent-based scientific applications
- Multimodal agentic systems integrating text, images, graphs, time series, and structured data
- Integration of LLMs with tools, memory, planning, and environment interaction
- Active learning strategies with LLMs in practical applications



This work is licensed under a Creative Commons Attribution 4.0 International License. *KDD '26, Jeju Island, Republic of Korea*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2259-2/2026/08
<https://doi.org/10.1145/3770855.3818253>

- Enhancing code generation for scientific practitioners
- Responsible, trustworthy, and ethical considerations of agent-based AI in society
- Benchmarks, datasets, and evaluation protocols for agentic AI and LLM-based agents
- Agentic workflows for AI-driven experimentation, hypothesis generation, and validation
- Self-evolving agents with autonomous learning, adaptation, and continual improvement

Target Audience. The target audience for the SciSoc Agents & LLM Workshop includes a diverse group of stakeholders from academia, industry, and government agencies who are at the forefront of artificial intelligence research and applications. Specifically, the workshop aims to attract computer scientists, data scientists, AI researchers, and domain experts in fields such as healthcare, environmental science, public policy, and social sciences who are interested in the integration and impact of Large Language Models and autonomous agents within their disciplines. Additionally, the event seeks to draw in policy makers and technology developers who are focused on ethical considerations and the practical deployment of AI technologies in real-world scenarios. This gathering will provide a unique opportunity for interdisciplinary exchange and collaboration, aiming to foster innovations that drive scientific discovery and address pivotal societal challenges through the advanced capabilities of LLMs and agentic AI systems.

Relevance to KDD. This workshop is highly relevant to KDD, as it focuses on data-driven advances that connect methodological innovation with real-world scientific and societal impact. LLMs and agentic AI systems are increasingly central to data mining and AI research, with strong interest across both academia and industry. Building on the momentum of LLM-focused events at recent KDD conferences, this workshop provides a dedicated forum for exploring how agentic AI can support scientific discovery and address societal challenges. By bringing together data scientists, AI researchers, domain experts, and policymakers, the workshop will foster interdisciplinary discussion on emerging methods, practical deployments, and open challenges. Through invited talks, panels, oral presentations, and posters, participants will exchange insights at the intersection of data mining, machine learning, science, and society—areas central to KDD’s mission.

Motivation & Rationale. Autonomous agents and LLMs are rapidly expanding beyond simple question-answering toward broader applications in science and society. Recent advances have shown their ability to solve complex reasoning tasks [3], support language translation [6], perform professional-level question answering [5], assist scientific discovery [9], and enable large-scale computational social science research [2]. These developments highlight the growing potential of agentic AI and LLMs to accelerate scientific progress and address societal challenges. The objective of the SciSoc Agents & LLMs Workshop at KDD’26 is to provide a forum for academic researchers and industry practitioners to present recent advances, real-world implementations, and emerging applications of LLM-powered agents for science and society. Given the current momentum in agentic AI and the interdisciplinary nature of KDD, this workshop will foster timely discussions and collaborations at the intersection of data mining, machine learning, scientific discovery, and societal impact. It will also complement broader LLM-focused

events at KDD by offering a specialized focus on scientific and societal advances.

2 Tentative Program Sketch

The half-day workshop is scheduled as a 3-hour event, excluding coffee and lunch breaks, as follows:

- The half-day session
 - 8:00 - 8:10 Welcome & Opening Remarks
 - 8:10 - 8:50 Keynote I Dr. James Zou
 - 8:50 - 9:20 Contributed Research Oral Talks I
 - 9:20 - 10:00 Keynote II Dr. Le Song
 - 10:00 - 10:40 Coffee break/Poster Session
 - 10:40 - 11:20 Keynote III Dr. Nitesh Chawla
 - 11:20 - 11:50 Contributed Research Oral Talks II
 - 11:50 - 12:00 Best Paper Award Ceremony & Final Remarks

Acknowledgment

B. Aditya Prakash was supported by the NSF under Award Numbers 1918770, 1955883, 2106961, 2403240, and 2200269, the NSF-Simons SkAI Institute, and the CDC MInD program. Yan Liu was supported by the NSF under Award Number 2425919.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint* (2023).
- [2] Gordon Dai, Weijia Zhang, Jinhao Li, Siqu Yang, Srihas Rao, Arthur Caetano, Misha Sra, et al. 2024. Artificial Leviathan: Exploring Social Evolution of LLM Agents Through the Lens of Hobbesian Social Contract Theory. *arXiv preprint arXiv:2406.14373* (2024).
- [3] Iddo Drori, Sarah Zhang, Reece Shuttlesworth, Leonard Tang, Albert Lu, Elizabeth Ke, Kevin Liu, Linda Chen, Sunny Tran, Newman Cheng, et al. 2022. A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. *Proceedings of the National Academy of Sciences* 119, 32 (2022), e2123433119.
- [4] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Anil Palepu, Petar Sirkovic, Artiom Myaskovsky, Felix Weissenberger, Keran Rong, Ryutaro Tanno, et al. 2025. Towards an AI co-scientist. *arXiv preprint arXiv:2502.18864* (2025).
- [5] Daniel Martin Katz, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 2024. Gpt-4 passes the bar exam. *Philosophical Transactions of the Royal Society A* 382, 2270 (2024), 20230254.
- [6] Tom Kocmi and Christian Federmann. 2023. Large language models are state-of-the-art evaluators of translation quality. *arXiv preprint arXiv:2302.14520* (2023).
- [7] Wanlu Lei, Caterina Fuster-Barceló, Gabriel Reder, Arrate Muñoz-Barrutia, and Wei Ouyang. 2024. BioImage. IO Chatbot: a community-driven AI assistant for integrative computational bioimaging. *nature methods* 21, 8 (2024), 1368–1370.
- [8] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Melissa Zhao, Aaron K Chow, Kenji Ikemura, Ahron Kim, Dimitra Pouli, Ankush Patel, et al. 2024. A multimodal generative AI copilot for human pathology. *Nature* 634, 8033 (2024), 466–473.
- [9] Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085* (2022).
- [10] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
- [11] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Hu. 2024. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data* 18, 6 (2024), 1–32.
- [12] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).